Received: 4 May 2022 | Revised: 26 June 2022 | Accepted: 28 June 2022

DOI: 10.1002/nbm.4795

RESEARCH ARTICLE

NMR IN BIOMEDICINE WILEY

Automated volumetric determination of high R_2^* regions in substantia nigra: A feasibility study of quantifying substantia nigra atrophy in progressive supranuclear palsy

Abel Worku Tessema[1,2] | Hansol Lee[1,3] | Yelim Gong[1] | Hwapyeong Cho[1] | Hamdia Murad Adem[2] | Ilwoo Lyu[4] | Jae-Hyeok Lee[5] | HyungJoon Cho[1]

[1]Department of Biomedical Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea

[2]School of Biomedical Engineering, Jimma Institute of Technology, Jimma University, Jimma, Ethiopia

[3]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital, Charlestown, MA, USA

[4]Department of Computer Science and Engineering, Ulsan National Institute of Science and Technology, Ulsan, South Korea

[5]Department of Neurology, Research Institute for Convergence of Biomedical Science and Technology, Pusan National University Yangsan Hospital, Yangsan, South Korea

Correspondence
HyungJoon Cho, Department of Biomedical Engineering, 105–222, 50, UNIST-gil, Eonyang-eup, Ulju-gun, Ulsan, South Korea.
Email: hjcho@unist.ac.kr

Jae-Hyeok Lee, Department of Neurology, 20, Geumo-ro, Mulgeum-eup, Yangsan-si, Gyeongsangnam-do, South Korea.
Email: jhlee.neuro@pusan.ac.kr

Ilwoo Lyu, Department of Computer Science and Engineering, 106, 801-5, 50, UNIST-gil, Eonyang-eup, Ulju-gun, Ulsan, South Korea.
Email: ilwoolyu@unist.ac.kr

Funding information
This work was partially supported by grants from the National Research Foundation of Korea from the Korean government (nos. 2018R1A6A1A03025810 and 2022R1A2C2011191). This research was also supported by the 2021 Joint Research Project of the Institute of Science and Technology. This study was also supported by a grant from the Korea Healthcare Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant no: HI14C1135).

The establishment of an unbiased protocol for the automated volumetric measurement of iron-rich regions in the substantia nigra (SN) is clinically important for diagnosing neurodegenerative diseases exhibiting midbrain atrophy, such as progressive supranuclear palsy (PSP). This study aimed to automatically quantify the volume and surface properties of the iron-rich 3D regions in the SN using the quantitative MRI-R_2^* map. Three hundred and sixty-seven slices of R_2^* map and susceptibility-weighted imaging (SWI) at 3-T MRI from healthy control (HC) individuals and Parkinson's disease (PD) patients were used to train customized U-net++ convolutional neural network based on expert-segmented masks. Age- and sex-matched participants were selected from HC, PD, and PSP groups to automate the volumetric determination of iron-rich areas in the SN. Dice similarity coefficient values between expert-segmented and detected masks from the proposed network were 0.91 ± 0.07 for R_2^* maps and 0.89 ± 0.08 for SWI. Reductions in iron-rich SN volume from the R_2^* map (SWI) were observed in PSP with area under the receiver operating characteristic curve values of 0.96 (0.89) and 0.98 (0.92) compared with HC and PD, respectively. The mean curvature of the PSP showed SN deformation along the side closer to the red nucleus. We demonstrated the automated volumetric measurement of iron-rich regions in the SN using deep learning can quantify the SN atrophy in PSP compared with PD and HC.

Abbreviations used: AC-PC, anterior commissure-posterior commissure; AI, artificial intelligence; AUC, area under the curve; BCE, binary cross-entropy; CLAHE, contrast-limited adaptive histogram equalization; DSC, Dice similarity coefficient; ELU, exponential linear unit; HC, healthy control; H-Y, Hoehn and Yahr; ICC, intraclass correlation coefficient; IOU, intersection over union; JSON, JavaScript object notation; MMSE, mini-mental state examination; MoCA, Montreal cognitive assessment; MRI, magnetic resonance imaging; PD, Parkinson's disease; PSP, progressive supranuclear palsy; QSM, quantitative susceptibility mapping; RN, red nucleus; ROC, receiver operating characteristic; ROI, region of interest; SN, substantia nigra; SWI, susceptibility-weighted imaging; TE, time of echo; TR, time of repetition; UPDRS, unified Parkinson's disease rating scale.

Abel Worku Tessema and Hansol Lee contributed equally to this work.

KEYWORDS
convolutional neural network, progressive supranuclear palsy, quantitative analysis, segmentation, substantia nigra

1 | INTRODUCTION

A high susceptibility difference due to the deposition of iron in the substantia nigra (SN) following disease progression causes an inhomogeneous magnetic field within the tissue and creates contrast on susceptibility-related MRI, enabling in vivo delineation of iron-rich SN areas.[1,2] Therefore, 3D SN volumes reconstructed from manually segmented masks on susceptibility-related MRI have often been investigated to evaluate morphological changes in iron-rich SNs of Parkinson's disease.[2–4] Accordingly, the volumetric measurements of iron-rich regions in the SN are important for diagnosing atypical Parkinsonian syndromes exhibiting midbrain atrophy, such as progressive supranuclear palsy (PSP) from Parkinson's disease (PD) patients and healthy subjects.[5,6] However, manual delineation of iron-rich SN volumes requires both anatomical knowledge and experience, is time-consuming, and can be rater-dependent.

Because unbiased and reproducible methods involving deep learning can potentially contribute to accurate segmentation, previous studies have investigated the automatic segmentation of SN using a convolutional neural network.[7–10] The segmentation of the SN has been mostly performed on qualitative MRI, such as neuromelanin-sensitive MRI and susceptibility-weighted imaging (SWI), whose contrasts may vary depending on the subjects, scanner, and scanning parameters.[7–9] In recent years, for reasons of necessity, segmentation of SN was also performed using one of quantitative MRI, quantitative susceptibility mapping (QSM).[10,11] To maximize the diagnostic efficacy of PSP through investigating morphological changes in the iron-rich SN region, a comparative study on the accuracy and diagnostic ability of automated measurements of SN volume and surface area between quantitative (R_2^* map) and qualitative (SWI) MRI is required, where both show iron-sensitive MR contrasts and are reconstructed from one gradient-based sequence.[12] Moreover, because the availability of clinical 3D SN image datasets for PSP is often limited, it is also important to test the feasibility of the diagnosis of PSP without including SN images of PSP while training the convolutional neural network.

In this work, we hypothesized that the volume of iron-rich regions in the SN of PSP brain is altered. We also hypothesized that an objective diagnosis through automatic segmentation of iron-rich SN using a convolutional neural network, based upon quantitative MRI-R_2^* maps, could superiorly contribute to the differentiation of PSP. Thus, this study comparatively assessed the performance of automatic segmentation of iron-rich SN using state-of-the-art and customized convolutional neural networks trained on R_2^* maps and SWI images. To investigate the clinical feasibility of unbiased diagnosis of PSP, the volume, surface area, and other surface properties were analyzed from automated iron-rich SN volumetric measurements on the age- and sex-matched HC, PD, and PSP brains.

2 | MATERIALS AND METHODS

The overall workflow of this study is shown in Figure 1. This study was approved by the Institutional Review Board of the Pusan National University Yangsan Hospital and the Ulsan National Institute of Science and Technology. Informed consent was obtained from all participants. All procedures were conducted in accordance with the Declaration of Helsinki.

2.1 | Study participants and inclusion criteria

A total of 93 participants were initially enrolled retrospectively, including 51 patients with PD, 11 patients with PSP, and 31 HC individuals. PD and PSP were diagnosed by a neurologist (J. H. L., with 15 years of experience in movement disorders) following established criteria.[13,14] Of the 93 participants, one patient with PD and one patient with PSP were excluded due to susceptibility artifacts in MRI. All participants underwent a mini-mental state examination (MMSE). The unified Parkinson's disease rating scale (UPDRS) part III and Hoehn and Yahr (H-Y) stages were tested to evaluate the disease severity in patients with PD and PSP.

2.2 | In vivo MRI

In vivo brain MRI was performed using a 3-T MRI scanner (MAGNETOM Skyra, Siemens) at Pusan National University Yangsan Hospital. 2D multiecho gradient-echo data were collected in parallel to the intercommissural (anterior commissure-posterior commissure [AC-PC]) line to

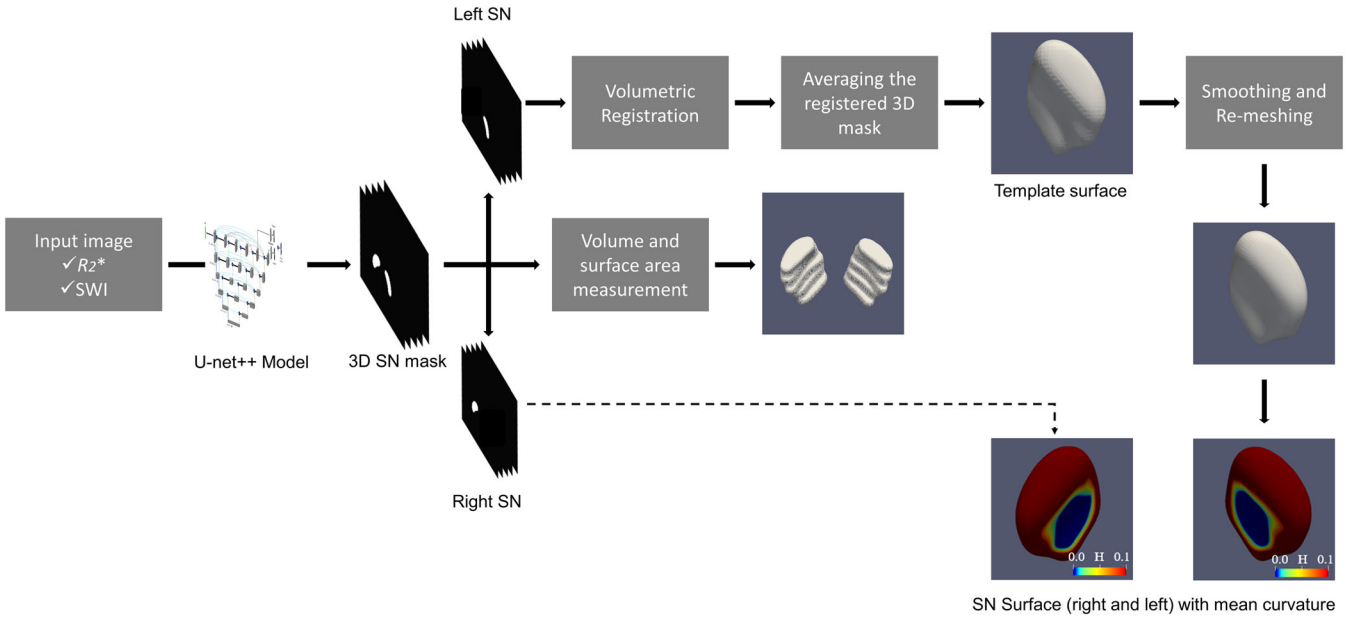

FIGURE 1 Block diagram of methodology. SN, substantia nigra; SWI, susceptibility-weighted imaging

encompass the whole brain for the reconstruction of the R_2^* map and SWI using the following parameters: time of repetition (TR) = 2030 ms; time of echo (TE) = 3.1–29.9 ms (ΔTE = 4.8, 5.5, ..., 5.5 ms); six echoes; matrix size = 192 × 192; voxel size = 1 × 1 × 2 mm; flip angle = 60°; distance factor = 10%; and number of slices = 60. The R_2^* value was calculated with the nonlinear fitting of a mono-exponential function of the R_2^* relaxation curve. SWI was generated by combining the magnitude and fourth power of the high pass-filtered positive phase mask of TE = 24.4 ms.[15]

2.3 | Data preparation

First, 10 individuals were selected from the HC and PD patient groups to match the cohort of PSP patients for age and sex distribution. Subsequently, five consecutive MRI slices per participant were used to test the segmentation algorithm and the quantitative analysis of the 3D SN structure. If the SN regions were included in six slices, the slice having the smallest and/or most challenging to notice SN region was removed in the test dataset to be consistent for all participants. A total of 367 MRI slices containing SN from the remaining 21 HC individuals and 40 patients with PD were used to train the networks. Because the SN is located in the midbrain, the original image was center-cropped to 64 × 64, followed by upsampling to 256 × 256 using the nearest interpolation to increase the size of the region of interest (ROI) and overall segmentation performance of the networks. The boundary of the SN was then delineated from the processed R_2^* map and SWI images independently, as the image contrast is different, by H. L. (6 years of experience in neuroimaging) using LabelMe (4.5.5), a software that creates a JavaScript Object Notation (JSON) file for labeled images, which was converted into the corresponding expert-segmented mask using an in-house Python (3.8, Python Software Foundation) program.[16] For the assessment of intra-rater reliability of manual SN segmentation, H. L. delineated the SN area again after a 7-month interval for 10 randomly selected participants (50 images of R_2^* map). For the inter-rater reliability, another rater (J. H. L.) delineated the SN area on 50 R_2^* maps as well. Data augmentation using ± 3-degree rotation, ± 90-degree rotation, and horizontal and vertical flipping, were applied to the training data of R_2^* and SWI. Contrast-limited adaptive histogram equalization (CLAHE) was additionally applied to SWI data to address the inconsistent image contrast.

2.4 | Convolutional neural networks

The convolutional neural networks based on U-net architecture with different designs of skip connections and application of deep supervision were selected and compared to segment iron-rich regions in the SN because of their high performances in segmenting objects of interest in the medical images using limited training data and ease of customization (Figure 2).[17,18] The basic architecture of U-net has an encoder or a feature extractor, a decoder, and skip connections that bring features from the encoder to the decoder with the same resolution for precise localization,

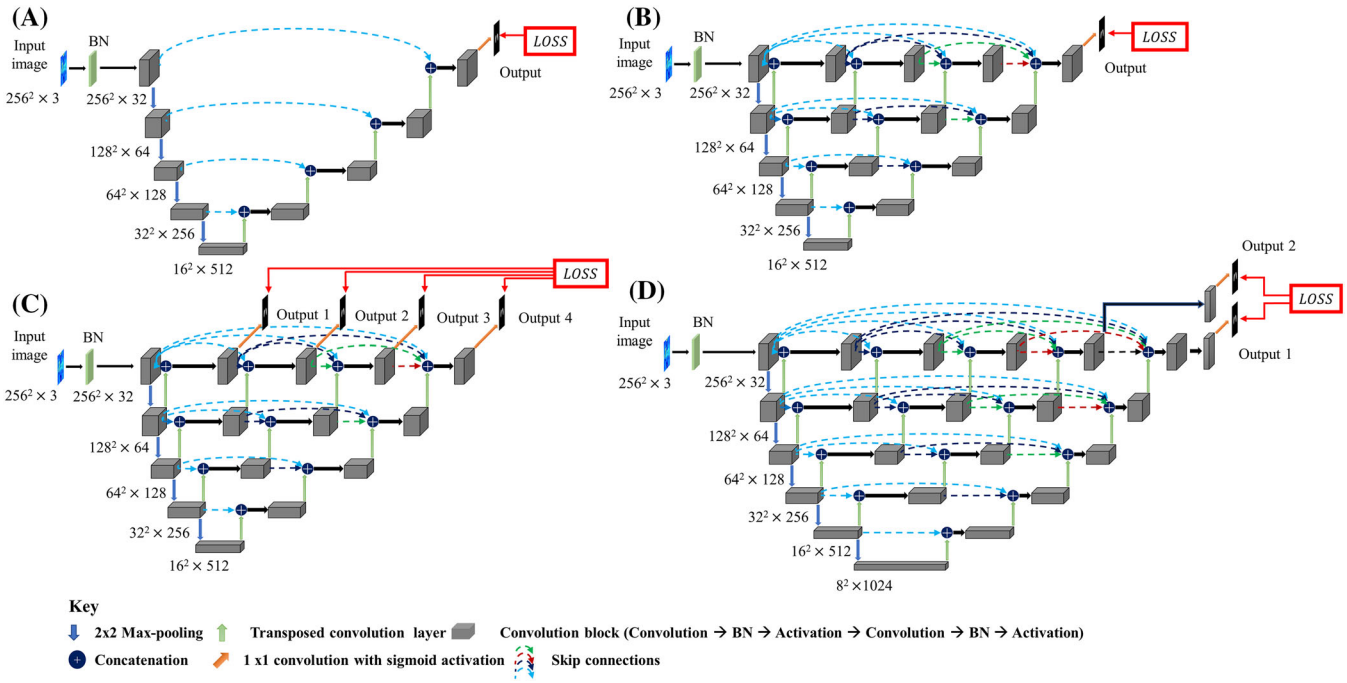

FIGURE 2 Neural network models used for segmentation of substantia nigra. (A) U-net. (B) Nested U-net/U-net++ without deep supervision. (C) U-net++ with deep supervision. (D) The proposed customized U-net++ with deep supervision. BN, batch normalization

as shown in Figure 2A.[17] Each block represents two convolution layers with a 3 × 3 kernel each followed by batch normalization and activation operations. In the encoder, the pooling layer is used after each convolution block to reduce the spatial dimension of the features while transposed convolution and concatenation layers precede each convolution block in the decoder. The skip connections in U-net combine low-level features from the encoder with high-level features from the decoder that are semantically different.

U-net++ improves the design of skip connections through the nesting of multiple U-net networks with different pyramid size, processing the low-level features from the encoder before it is fused to semantically rich features from the decoder.[18] In addition, the loss is backpropagated to all nested networks simultaneously through deep supervision. In this study, a customized U-net++ model was proposed by applying the deep supervision only to the last two deepest network pyramids. A dropout layer was added in the middle of each convolutional block for the additional regularization. The pyramid of the network was increased by one step making the customized network deeper than other networks. Lastly, an additional convolutional block with a feature depth of 16 was inserted before applying a 1 × 1 convolution with sigmoid activation (Equations 1 and 2). Modification of the architecture and fine tuning of the network hyperparameters were performed to improve the performance of the networks in segmenting the iron-rich regions of SN. An exponential linear unit (ELU) with an alpha of 1.5 (Equations 3 and 4), AdaMax optimizer with adaptive learning rate (initial learning rate = 10^{-5}), L_2-kernel regularization ($L_2 = 10^{-2}$), and "he-uniform" kernel initializer were selected after a rigorous assessment of segmentation performance.

$$Sigmoid(x) = \frac{1}{1+e^{-x}} \tag{1}$$

$$Sigmoid'(x) = \frac{e^{-x}}{(e^{-x}+1)^2} \tag{2}$$

$$ELU(x) = \begin{cases} x, & x \geq 0 \\ \alpha(e^x - 1), & otherwise \end{cases} \tag{3}$$

$$ELU'(x) = \begin{cases} 1, & x \geq 0 \\ ELU(x) + \alpha, & otherwise \end{cases} \tag{4}$$

where x is input and α is a parameter used to saturate the negative input.

Dropout rates of 0.3 and 0.2 were also used to train the proposed network on R_2^* and SWI data, respectively. All networks were trained on three Nvidia GeForce RTX 3090 (Nvidia Corporation, Santa Clara, CA, USA) graphics processors with a batch size of 48 using data parallelization.

The maximum number of epochs was set to 500 and a stopping criterion was implemented by tracking the validation loss calculated using 10% of the training data selected randomly to stop the training before overfitting occurred. Binary cross-entropy (BCE) loss was exploited as a loss function to train the networks according to Equation 5. The models were implemented using TensorFlow (2.5, Google Brain Team, Google LLC.) installed in Python (3.8, Python Software Foundation).

$$\begin{aligned}\text{BCE loss} &= -\frac{1}{N}\sum_{i=1}^{N} y_i * \log(P_i) + (1-y_i) * \log(1-P_i) \\ &= -\frac{1}{N}\sum_{i=1}^{N} y_i * \log(P_i)\end{aligned}, \tag{5}$$

where P_i is the probability of foreground and 1 – P_i is the probability of background.

2.5 | Volumetric and shape analysis

A morphological opening was applied to the detected mask to remove the scattered pixels. After compensating for the distance factor between each slice, a 3D SN mask was generated for each participant by cascading the detected mask of five consecutive slices. The SN volume was then measured by counting the number of voxels in the 3D mask. The surface area was estimated on terraced field areas of the concatenated 3D mask after compensating for the spatial resolution, slice thickness, and distance factor using an in-house MATLAB code (R2020b, MathWorks).

The template of the 3D SN volume for each group was produced separately for the right and left SNs using volumetric registration to align all volumes and calculate the average. The final template SN surface was generated from the template volume, smoothed, and re-tessellated using standard icosahedral re-tessellation to reconstruct the SN surface with consistent tessellation.[19–21] The mean curvature and principal curvatures were extracted from the re-tessellated template surface with 2562 vertices for each hemisphere to explore changes in the shape of the SN. At each vertex, the shape index is calculated with $(\frac{2}{\pi}\tan^{-1}\frac{K1+K2}{K1-K2})$,[22] where $K1$ and $K2$ are the maximum and minimum principal curvatures, respectively.

2.6 | Statistical analysis

Demographic and clinical data were compared between the three groups of testing data using the chi-square test, Kruskal–Wallis H test, and Mann–Whitney U test. A Bonferroni post hoc correction using the Mann–Whitney U test was performed for three comparisons between groups with a statistical significance threshold of p less than 0.0167 (0.05/3). The intraclass correlation coefficient (ICC) was calculated to determine the intra-rater and inter-rater reliability between two segmentation products with a 7-month interval from one rater and between two segmentation products from two raters on the R_2^* map. The Dice similarity coefficient (DSC) and intersection over union (IOU) were measured between the expert-segmented mask and the detected mask to evaluate and compare the segmentation performance of the networks on the testing dataset according to Equations 6 and (7). The volume and surface area of 10 participants in each group generated from the detected mask were compared between the three groups using the Kruskal–Wallis test followed by Bonferroni post hoc correction. The mean curvature and shape index of the 3D SN template surface with 5124 vertices corresponding to the surfaces of the SN in both hemispheres for each group were compared between groups. The vertices that formed the concave regions of the SN were also compared. The correlation between the volume and surface area was tested using linear regression and Pearson's correlation coefficient. Receiver operating characteristic (ROC) analysis was also performed to calculate the area under the ROC curve (AUC) to assess the overall diagnostic performance to distinguish between groups of patients and HC individuals. Statistical significance was set at p less than 0.05.

$$\text{DSC}(y^t, y^p) = \frac{2 * \sum_{i=1}^{r}\sum_{j=1}^{c}\left(y_{ij}^t * y_{ij}^p\right)}{\sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}^t + \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}^p} = \frac{2\,(y^t \cap y^p)}{y^t + y^p} \tag{6}$$

$$\text{IOU}(y^t, y^p) = \frac{\sum_{i=1}^{r}\sum_{j=1}^{c}\left(y_{ij}^t * y_{ij}^p\right)}{\sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}^t + \sum_{i=1}^{r}\sum_{j=1}^{c} y_{ij}^p - \sum_{i=1}^{r}\sum_{j=1}^{c}\left(y_{ij}^t * y_{ij}^p\right)} = \frac{y^t \cap y^p}{y^t \cup y^p}, \tag{7}$$

where y^t is a ground truth image, y^p is a predicted image, and r and c are the height and width of the image, respectively.

3 | RESULTS

Demographic information for each group is summarized in Table 1. There was no difference in sex and age between groups ($p > 0.05$). However, due to the greater disease severity of PSP patients, the distributions of MMSE, UPDRS III, and H-Y stage were different between the PD and PSP groups (all $p < 0.05$). The ICC values for the intra-rater and inter-rater reliability of the SN segmentations using the R_2^* map were 0.997 and 0.987, respectively, showing almost perfect agreement based on the Landis and Koch criteria.[23]

3.1 | SN segmentation

The comparison of computational cost, inference time, and segmentation performance of different networks on testing data is shown in Table 2. The customized U-net++ network outperformed all other models in terms of the segmentation performance using both R_2^* map and SWI data with costs of computational complexity, training, and inference time. The performance of the proposed model on testing images from R_2^* map and SWI is shown for both the rostral and caudal levels of the three groups in Figures 3 and S1, respectively. The overlap of the expert-segmented mask with the detected mask is presented with a corresponding original image for each group, where red indicates true positives, green indicates false negatives, and blue indicates false positives. The DSC and IOU values between the expert-segmented mask and detected mask were 0.91 ± 0.07 (standard deviation) and 0.83 ± 0.10, respectively, for R_2^* data at a confidence threshold of 0.8 on 150 test images. DSC and IOU values for SWI data were lower than R_2^* data with 0.89 ± 0.08 and 0.81 ± 0.11, respectively, at a confidence threshold of 0.6.

TABLE 1 Clinical and demographic characteristics of PD and PSP patients along with HC participants

					Post hoc (*p*)		
	HC (N = 10)	PD (N = 10)	PSP (N = 10)	Group comparison (*p*)	HC versus PD	HC versus PSP	PD versus PSP
Gender (M/F)	6/4	6/4	6/4	1[a]	1	1	1
Age (years)	63.3 ± 2.2	63.9 ± 2.0	65.8 ± 3.2	0.09[b]	0.51	0.07	0.08
MMSE	28.6 ± 1.5	27.3 ± 2.8	23.4 ± 4.1	**0.01**[b]*	0.48	**0.003****	**0.03***
UPDRS III	-	24.5 ± 5.7	37.8 ± 11.5	**0.005**[c]**	-	-	**0.005****
H-Y stage	-	2.5 ± 0.4	3.7 ± 0.9	**0.001**[c]**	-	-	**0.001****
Duration (years)	-	6.8 ± 3.7	4.4 ± 2.1	**0.04**[c]*	-	-	**0.04***

Data are shown as mean ± standard deviation.
Abbreviations: HC, healthy controls; H-Y, Hoehn and Yahr; MMSE, mini-mental state examination; PD, Parkinson's disease; PSP, progressive supranuclear palsy; UPDRS III, the motor part of the unified Parkinson's disease rating scale. The significant *p* values are displayed in bold.
Post hoc (*p*): *p* values from post hoc analysis of χ^2 test[a] and Mann–Whitney U test[b,c].
*$p < 0.05$ and **$p < 0.005$. (For post hoc analysis: *$p < 0.05$ and **$p < 0.005$ after Bonferroni correction.)
[a]χ^2 test.
[b]Kruskal–Wallis H test.
[c]Mann–Whitney U test.

TABLE 2 Comparison among convolutional neural networks

	Network properties			Segmentation performance	
Model	Number of parameters (M)	Training time (h)	Inference time (s)	Mean IOU	Mean DSC
U-net	~**7.8**	~**3**	~**0.025**	(0.74, 0.73)	(0.84, 0.83)
U-net++	~9	~7	~0.035	(0.75, 0.74)	(0.86, 0.85)
U-net++ with DS	~9	~7.5	~0.035	(0.76, 0.75)	(0.87, 0.85)
Customized U-net++	~36	~10	~0.040	**(0.83, 0.81)**	**(0.91, 0.89)**

The number of parameters refers to all trainable and nontrainable parameters. The first and second numbers in parenthesis show the results on R_2^* and SWI data, respectively.
Abbreviations: DS, deep supervision; DSC, Dice similarity coefficient; Inference time, time to process a single image on a single graphics processing unit; IOU, intersection over union; M, million; SWI, susceptibility-weighted imaging. The numbers in bold shows maximum performance.

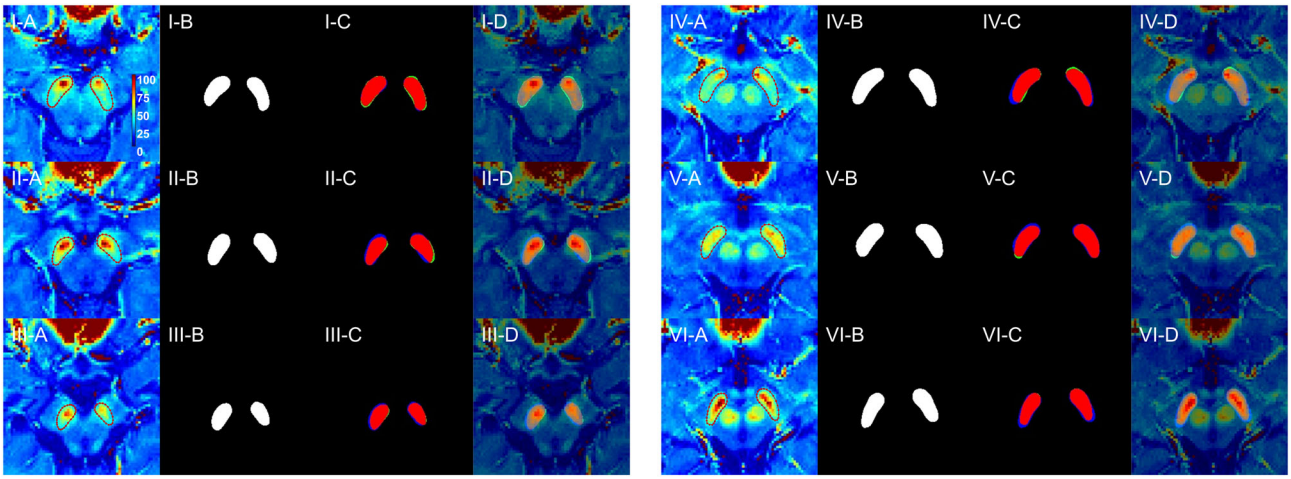

FIGURE 3 Detection of SN in R_2^* maps using the U-net++ model. I-III show the caudal level. IV-VI show the rostral level. I and IV are representative of HC individuals (61-year-old male). II and V are representative of PD patients (61-year-old male). III and VI are representative of PSP patients (60-year-old male). (A) R_2^* maps with the expert-segmented mask of iron-rich SN. (B) Detected mask from U-net++. (C) The overlap of expert-segmented mask and detected mask. (D) The overlap of R_2^* maps with performance masks. The red area in (C) shows the true positives, the green area shows the false negatives, and the blue area shows the false positives. HC, healthy control; PD, Parkinson's disease; PSP, progressive supranuclear palsy; SN, substantia nigra

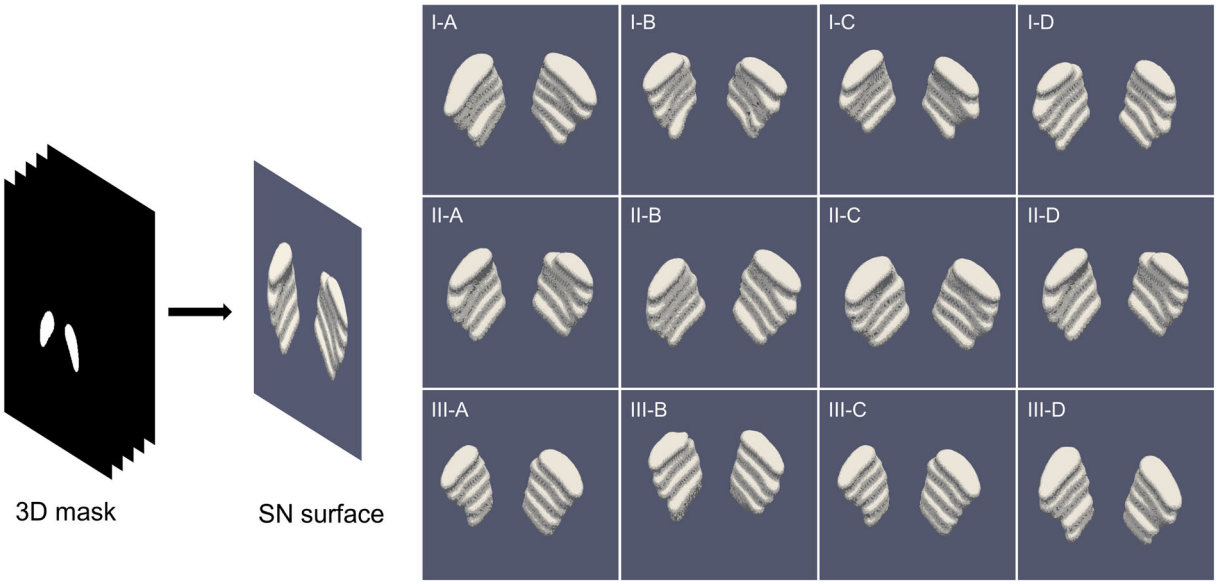

FIGURE 4 SN surfaces constructed from the generated mask using R_2^* maps. I represents HC individuals. II represents PD patients. III represents PSP patients. (A-D) Four sample SN surfaces randomly selected in each group. HC, healthy control; PD, Parkinson's disease; PSP, progressive supranuclear palsy; SN, substantia nigra

3.2 | Quantitative analysis

3D SN surfaces generated from five consecutive 2D detected masks are shown from four individuals in each group (Figures 4 and S2). The SN volume ($R_2^* = 1128.2 \pm 232.2$ mm^3; SWI $= 1206.9 \pm 193.8$ mm^3) in PSP was reduced compared with HC and PD ($R_2^* = 1474.7 \pm 124.4$ mm^3 with $p < 0.001$ in HC, 1445.3 ± 64.4 mm^3 with $p < 0.001$ in PD; SWI $= 1527.3 \pm 179.4$ mm^3 with $p < 0.004$ in HC, 1506.2 ± 79.8 mm^3 with $p < 0.002$ in PD) (Table 3). The surface area consistently showed a significant difference in PSP ($p < 0.05$) compared with the HC and PD groups. However, there was no difference between the HC and PD groups in either case ($p > 0.05$).

The SN volume and surface area were highly associated with each other in the three groups (Figure 5). Pearson correlation values were 0.92 (HC), 0.77 (PD), 0.96 (PSP), and 0.96 (all) for R_2^* and 0.93 (HC), 0.76 (PD), 0.90 (PSP), and 0.93 (all) for SWI. ROC curve analysis indicated that the

TABLE 3 SN volume and surface area of PD and PSP patients along with HC participants

	HC	PD	PSP	Kruskal–Wallis		Post hoc (*p*)		
				χ^2	p	HC versus PD	HC versus PSP	PD versus PSP
SN volume (mm^3)								
R_2^*	1474.7 ± 124.4	1445.3 ± 64.4	1128.2 ± 232.2	17.36	**< 0.001****	0.43	**< 0.001****	**< 0.001****
SWI	1527.3 ± 179.4	1506.2 ± 79.8	1206.9 ± 193.8	13.17	**0.001****	0.27	**0.004****	**0.002****
SN surface area (mm^2)								
R_2^*	1564.0 ± 84.7	1569.2 ± 49.2	1357.9 ± 176.7	14.17	**<0.001****	0.68	**0.003****	**0.001****
SWI	1602.8 ± 106.9	1615.6 ± 58.2	1406.9 ± 171.1	9.55	**0.008***	0.91	0.017	**0.005***

Data are shown as mean ± standard deviation. The measurements were the sum of left and right SNs.
Post hoc (*p*): *p* values from post hoc analysis of Mann–Whitney U test. *$p < 0.05$ and **$p < 0.005$. (For post hoc analysis: *$p < 0.05$ and **$p < 0.005$ after Bonferroni correction). The significant *p* values are displayed in bold.
Abbreviations: HC, healthy control; PD, Parkinson's disease; PSP, progressive supranuclear palsy; SN, substantia nigra; SWI, susceptibility-weighted imaging.

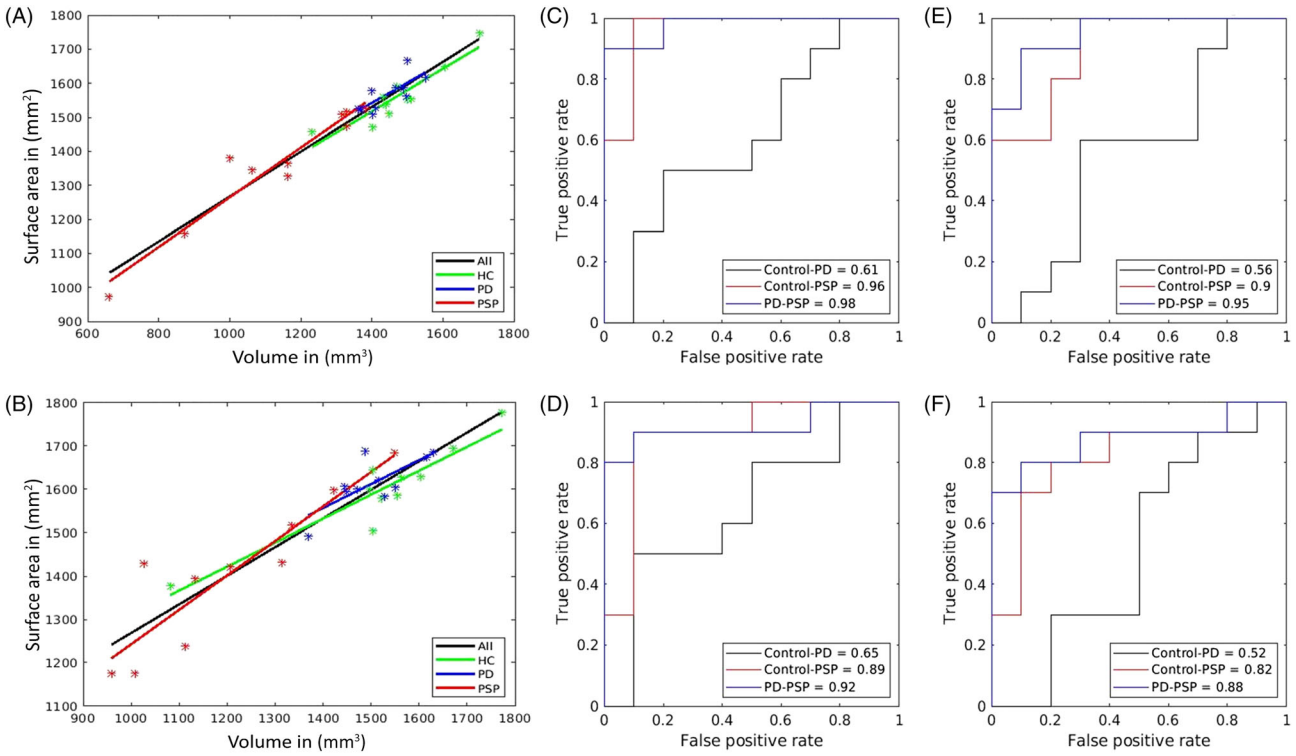

FIGURE 5 Correlation between SN volume and surface area and ROC curve analysis. (A, B) The correlation between SN volume and surface area using R_2^* and SWI data, respectively. (C, D) ROC curves of SN volume from R_2^* and SWI data, respectively. (E, F) ROC of the SN surface area from R_2^* and SWI data, respectively. HC, healthy control; PD, Parkinson's disease; PSP, progressive supranuclear palsy; ROC, receiver operating characteristic; SN, substantia nigra; SWI, susceptibility-weighted imaging

AUC values of the SN volume were high enough to distinguish PSP from the other two groups ($R_2^* = 0.96$ with HC, 0.98 with PD; SWI = 0.89 with HC, 0.92 with PD). However, AUC values were low for the differentiation between HC and PD patients ($R_2^* = 0.61$; SWI = 0.65 with HC). These results were consistent with the analysis of the SN surface area with AUC values of 0.9 between HC and PSP, 0.95 between PD and PSP, and 0.56 between HC and PD for R_2^*, and 0.82 between HC and PSP, 0.88 between PD and PSP, and 0.52 between HC and PD for SWI.

Template SN surfaces with mean curvature properties and the distribution of vertices with negative mean curvatures are presented for each group in Figures 6 and S3. The SN volume shrinkage and the change in the negative mean curvature were clearly observed in PSP. Table 4 shows that the negative mean curvature was shifted to the positive side in PSP compared with HC and PD. The mean curvature was significantly different in PSP (0.20 ± 0.14) only when using the SN surface acquired from SWI data, compared with HC (0.19 ± 0.16 [$p < 0.001$]) and PD (0.19 ±

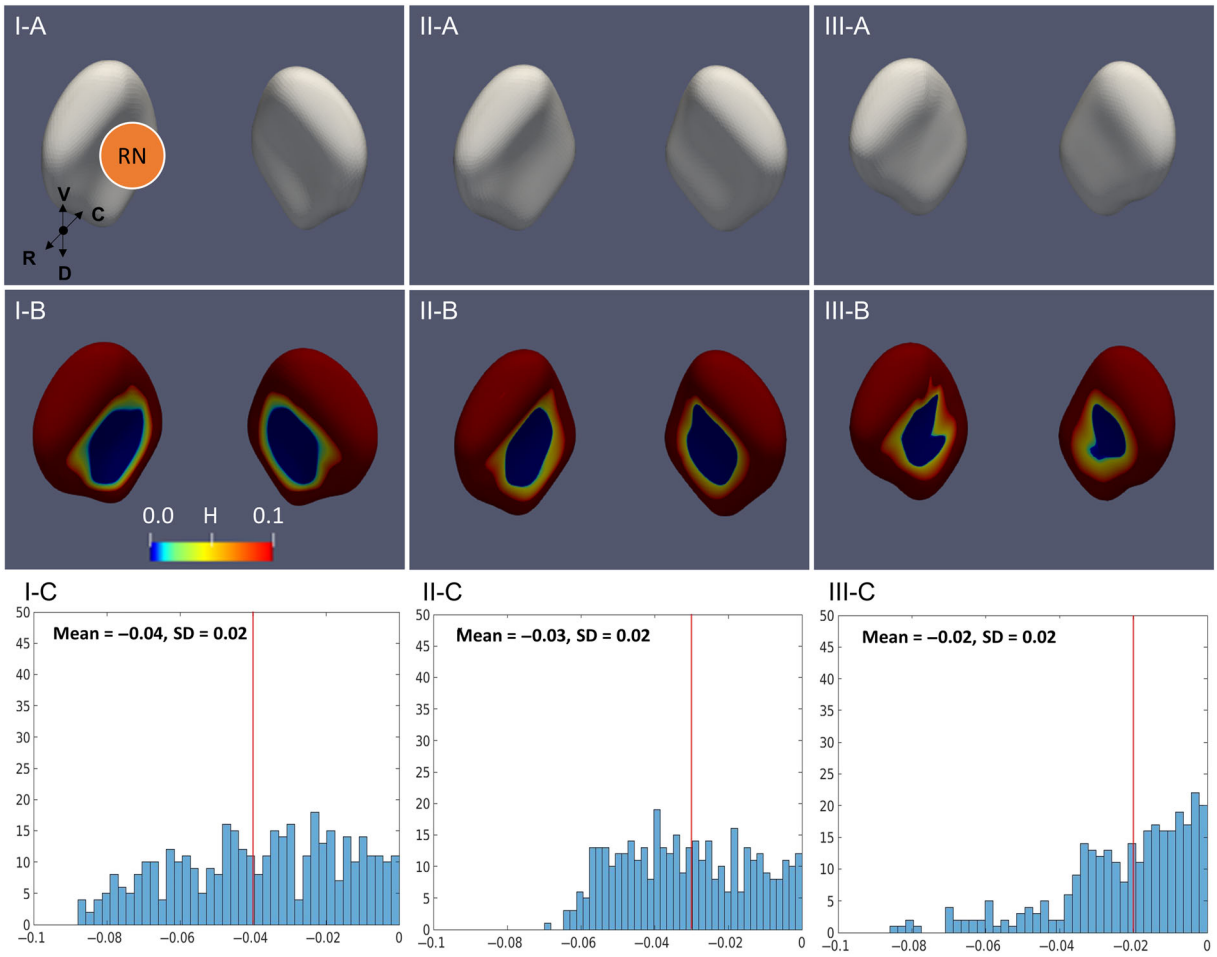

FIGURE 6 The representative template surface and curvature properties of the SN surface from R_2^* maps. I represents HCs. II represents PD patients. III represents PSP patients. (A) The template surfaces. (B) The template surfaces with mean curvature scaled between 0 and 0.1 mm^{-1}. (C) Histograms of the negative mean curvature of SN surfaces. H in the scale bar is the mean curvature property. C, caudal; D, dorsal; HC, healthy control; PD, Parkinson's disease; PSP, progressive supranuclear palsy; R, rostral; RN, red nucleus; SN, substantia nigra; V, ventral

TABLE 4 Mean curvature of PD and PSP patients along with HC participants

				Kruskal–Wallis		Post hoc (*p*)		
	HC	PD	PSP	χ^2	p	HC versus PD	HC versus PSP	PD versus PSP
Mean curvature (mm^{-1})								
R_2^*	0.19 ± 0.15	0.19 ± 0.15	0.19 ± 0.15	3.85	0.15	0.42	0.06	0.21
SWI	0.19 ± 0.16	0.19 ± 0.15	0.20 ± 0.14	18.06	**< 0.001****	0.61	**< 0.001****	**< 0.001****
Negative mean curvature (mm^{-1})								
R_2^*	−0.04 ± 0.02	−0.03 ± 0.02	−0.02 ± 0.02	91.97	**< 0.001****	**< 0.001****	**< 0.001****	**< 0.001****
SWI	−0.05 ± 0.03	−0.04 ± 0.02	−0.02 ± 0.01	186.91	**< 0.001****	**< 0.001****	**< 0.001****	**< 0.001****

Data are shown as mean ± standard deviation.
Post hoc (*p*): *p* values from post hoc analysis of Mann–Whitney U test. **$p < 0.005$. (For post hoc analysis: **$p < 0.005$ after Bonferroni correction). The significant *p* values are displayed in bold. Abbreviations: HC, healthy control; PD, Parkinson's disease; PSP, progressive supranuclear palsy; SWI, susceptibility-weighted imaging.

0.15 [$p < 0.001$]). However, there was no difference between surfaces from R_2^* data (HC = 0.19 ± 0.15; PD = 0.19 ± 0.15; and PSP = 0.19 ± 0.15). The negative mean curvature was also significantly different in PSP (R_2^*: −0.02 ± 0.02; SWI: −0.02 ± 0.01) than in HC (R_2^*: HC = −0.04 ± 0.02 [$p < 0.001$] and PD = −0.03 ± 0.02 [$p < 0.001$]; SWI: HC = −0.05 ± 0.03 [$p < 0.001$] and PD = −0.04 ± 0.02 [$p < 0.001$]). Although there was no difference in the mean curvature between HC and PD (R_2^*: $p = 0.70$; SWI: $p = 0.80$), PD was differentiated from HC by the negative mean

curvature (R_2^*: $p < 0.001$; SWI: $p < 0.001$). The shape index of each group is shown in Figure S4. Similar to the mean curvature analysis, the concave areas of the SN that enfold the red nucleus (RN) shrunk in PSP. However, it was not possible to distinguish PD and PSP patients from HC individuals (all $p > 0.05$).

4 | DISCUSSION

We examined the hypothesis that the SN structure is specifically altered with significant volume loss in patients with PSP. 3D iron-rich regions in the SN tissue can be reconstructed after automatic segmentation using a customized U-net++ convolutional neural network, which was trained based on expert-segmented SN masks of HC and PD using the quantitative MRI-R_2^* maps. The automated measurements of volume, surface area, and mean curvature of the iron-rich region in the SN allowed differentiation of patients with PSP from HC and patients with PD.[24] Therefore, we demonstrated the unbiased and reproducible quantification of the SN atrophy in the patients with PSP objectively through automatic segmentation of iron-rich regions in the SN using deep learning.

An automatic system that provides volume, surface area, and other shape properties of 3D iron-rich regions in SN from raw MRI data using deep learning is highly valuable because it contributes to computer-aided diagnosis that can reduce subjectivity and increase overall sensitivity in stratifying patients. Although previous studies with convolutional neural networks have been proposed for automatic segmentation of SN, most of them focused on qualitative MRI imaging.[7,8,25] In this study, customized U-net++ was implemented and a comparative study was performed with other U-net types to assess the segmentation performance using both qualitative SWI and quantitative R_2^* map. The 3D SN volume and surface were subsequently reconstructed using the customized network. The proposed model outperformed others with a larger margin in terms of segmenting the iron-rich regions in the SN. The application of deep supervision to all nested networks of U-net++ with different pyramid depths is prone to underfitting and overfitting of shallowest and deepest networks, respectively, which is a challenging problem to solve simultaneously. Applying the deep supervision to the deepest pyramids only and using dropout and batch normalization together prevented overfitting and improved the overall performance of the network. Using the proposed network, the DSC between the expert-segmented mask and the detected mask was 0.91 in the segmentation of SN from the R_2^* maps. Although the inconsistent contrast in SWI was compensated by CLAHE and an optimized convolutional neural network, the DSC for SWI was 0.89, which is slightly lower than that of the R_2^* maps. However, the statistics remained approximately the same in both cases, which is important in terms of clinical application, as SWI is commonly used for the diagnosis of various brain disorders. Although R_2^* maps are quantitative and sensitive to iron deposition around the SN, which is helpful for the automatic SN segmentation, slightly longer acquisition time compared with SWI should be cautiously considered to reduce motion artifacts from the patients.

A few prior studies have been conducted to correctly segment the SN region using different methods involving both quantitative and qualitative MRI. We compared the performance of SN segmentation in our study with that of previous studies. One recent study reported the segmentation of SN using QSM data by advanced image analysis with DSC values of 0.75 and 0.76 in left and right SN, respectively.[11] Different neuromelanin- and iron-containing midbrain nuclei were also segmented using a template of the brain from QSM data and achieved a DSC of 0.85 and 0.87 for neuromelanin- and iron-containing SN, respectively.[10] 3D SN segmentation on neuromelanin-sensitive MRI and SWI using a V-net convolutional neural network yielded an average DSC in five-fold cross-validation of 0.7.[8] Further, SN segmentation was also performed using an ensemble of five state-of-the-art convolutional neural networks that achieved average DSC in five-fold cross-validation of 0.87–0.93 using SWI images collected only from healthy individuals, which may be highly computationally intensive.[7] These comparisons validated that our customized U-net++ network consistently produced efficient SN segmentation results for HC, PD, and PSP compared with other methods.

Midbrain shrinkage, which is a marked characteristic of patients with PSP, is observed in conventional structural MRI by visual inspection.[26,27] In this study, it was demonstrated that PSP, which has severe neuronal loss and heavy iron accumulation in the SN, can be differentiated from HC and PD through quantification of SN volume, surface area, and shape properties following deep learning-based automatic SN segmentation. The distribution of mean curvature values clearly showed deformation of the SN surface in PSP due to the severe atrophy and flattened surface on the side of the enfolding RN, which may also be induced by iron accumulation along with the myelinated fibers of the third cranial nerve (oculomotor nerve) passing through the gap between the SN and RN.[27] Progression of PD may also cause morphological changes in the SN. There was no difference in the volume and surface area between HC and PD patients; however, because various factors including age, disease duration, and disease severity of PD patients that affect the SN morphological changes were not considered in our cohort, negative mean curvature values provided a possibility of differentiating between PD patients and HC individuals.

In previous studies, there was high inconsistency in SN volume measurements using MRI data. Variability in the MRI protocols, the geometry of slices, and the segmentation mechanisms of SN are among the main reasons for nonreproducible SN volume measurements. Because SN is very small in whole-brain structural MRI and each MRI contrast shows a disparate boundary, accurate SN segmentation is challenging and remains controversial.[2,28–30] In most previous studies, measured bilateral SN volumes range from 110 to 700 mm^3, showing a high degree of variation.[3,10,30,31] In one previous paper investigating susceptibility-related MRI, average SN volumes of 541 and 621 mm^3 for HC and PD, respectively, were reported from manually segmented masks surrounding the hypointense area in T_2^*-weighted images.[3] By comparison, the bilateral SN volumes in our study ranges from 1128 to 1527 mm^3. Therefore, the measured SN volume of our study was several times higher than that of

others. Our in vivo SN volume measurement was quantitatively supported by applying the same procedures in postmortem R_2^* maps of ex vivo midbrain tissues (see the supporting information: Methods, Results, Table S1, and Figure S5).[32]

In this study, the expert-segmented SN masks were delineated in the R_2^* map and SWI separately, which may not reveal the precise SN boundary, but are sensitive to the area of iron accumulation that creates a large magnetic susceptibility difference within and around the SN. Although the measured SN volume using iron-sensitive MRI images may be larger than the actual histological SN boundary, because iron can also accumulate outside the SN, 3D SN volume related to iron deposits was reconstructed, which helps to differentiate PSP patients from other groups.[33] The accurate in vivo segmentation of SN boundary with MRI will be needed to investigate the distinct morphological change of SN tissue with the disease progression.

This study had several limitations. First, the training datasets in this study purposedly included HC and PD. The DSC values were 0.92 for HC, 0.92 for PD, and 0.88 for PSP using R_2^* maps and 0.90 for HC, 0.90 for PD, and 0.87 for PSP using SWI, showing a slightly lower performance in segmenting the SN of PSP, which likely derives from the absence of PSP SN images in the training datasets. On the other hand, it is important to note that automatic PSP diagnosis was still feasible with high AUC values, even without PSP in the training dataset. Second, the extent and direction of anisotropic deformation of the SN in the PSP brains should be further investigated in the 3D axis with higher resolution images, although mean curvature values showed an altered morphology in PSP relative to HC individuals.

In conclusion, this study highlighted the automatic reconstruction of 3D iron-rich SN volumes from the segmentation in susceptibility-related MRI, specifically quantitative R_2^* maps, using deep learning to help in the diagnosis of diseased brains. We demonstrated the automated volumetric measurement of iron-rich regions in the SN using deep learning can quantify the SN atrophy in PSP compared with PD and HC.

ACKNOWLEDGMENTS

This work was a collaborative study between the Ulsan National Institute of Science and Technology and the Pusan National University Yangsan Hospital. This work was partially supported by grants from the National Research Foundation of Korea from the Korean government (nos. 2018R1A6A1A03025810 and 2022R1A2C2011191). This research was also supported by the 2021 Joint Research Project of the Institute of Science and Technology. This study was also supported by a grant from the Korea Healthcare Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant no: HI14C1135).

CONFLICTS OF INTEREST

The authors declare that they have no competing financial interests or personal relationships that could have influenced the research in this study.

REFERENCES

1. Hametner S, Endmayr V, Deistung A, et al. The influence of brain iron and myelin on magnetic susceptibility and effective transverse relaxation-A biochemical and histological validation study. *NeuroImage*. 2018;179:117-133.
2. Huddleston DE, Langley J, Dusek P, et al. Imaging Parkinsonian pathology in substantia nigra with MRI. *Curr Radiol Rep*. 2018;6(4):1-13.
3. Kwon DH, Kim JM, Oh SH, et al. Seven-Tesla magnetic resonance images of the substantia nigra in Parkinson disease. *Ann Neurol*. 2012;71(2):267-277. doi:10.1002/ana.22592
4. Lehéricy S, Sharman MA, Santos CLD, Paquin R, Gallea C. Magnetic resonance imaging of the substantia nigra in Parkinson's disease. *Mov Disord*. 2012;27(7):822-830.
5. Massey LA, Micallef C, Paviour DC, et al. Conventional magnetic resonance imaging in confirmed progressive supranuclear palsy and multiple system atrophy. *Mov Disord*. 2012;27(14):1754-1762. doi:10.1002/mds.24968
6. Taniguchi D, Hatano T, Kamagata K, et al. Neuromelanin imaging and midbrain volumetry in progressive supranuclear palsy and Parkinson's disease. *Mov Disord*. 2018;33(9):1488-1492. doi:10.1002/mds.27365
7. Beliveau V, Nørgaard M, Birkl C, Seppi K, Scherfler C. Automated segmentation of deep brain nuclei using convolutional neural networks and susceptibility weighted imaging. *Hum Brain Mapp*. 2021;42(15):4809-4822. doi:10.1002/hbm.25604
8. Kang J, Kim H, Kim E, et al. Convolutional neural network-based automatic segmentation of substantia nigra on nigrosome and neuromelanin sensitive MR images. *Investig Magn Reason Imaging*. 2021;25(3):156-163. doi:10.13104/imri.2021.25.3.156
9. Gaurav R, Pyatigorskaya N, Biondetti E, et al. Deep learning-based neuromelanin MRI changes of isolated REM sleep behavior disorder. *Mov Disord*. 2022;37(5):1064-1069. doi:10.1002/mds.28933
10. Jin Z, Wang Y, Jokar M, et al. Automatic detection of neuromelanin and iron in the midbrain nuclei using a magnetic resonance imaging-based brain template. *Hum Brain Mapp*. 2022;43(6):2011-2025. doi:10.1002/hbm.25770
11. Basukala D, Mukundan R, Lim A, et al. Automated segmentation of substantia nigra and red nucleus using quantitative susceptibility mapping images: Application to Parkinson's disease. *Comput Electr Eng*. 2021;91:107091. doi:10.1016/j.compeleceng.2021.107091
12. Lee JH, Lee MS. Brain iron accumulation in atypical Parkinsonian syndromes: in vivo MRI evidences for distinctive patterns. *Front Neurol*. 2019;10:74. doi:10.3389/fneur.2019.00074
13. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiat*. 1992;55(3):181-184. doi:10.1136/jnnp.55.3.181
14. Höglinger GU, Respondek G, Stamelou M, et al. Clinical diagnosis of progressive supranuclear palsy: the movement disorder society criteria. *Mov Disord*. 2017;32(6):853-864. doi:10.1002/mds.26987

15. Haacke EM, Xu Y, Cheng YCN, Reichenbach JR. Susceptibility weighted imaging (SWI). *Magn Reason Med*. 2004;52(3):612-618. doi:10.1002/mrm.20198
16. Wada K. Labelme: Image polygonal annotation with python v4.5.5. 2016. https://github.com/wkentaro/labelme
17. Ronneberger O, Fischer P, Brox T. October. U-net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical image computing and computer-assisted intervention*. Springer; 2015:234-241.
18. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. In: *Deep learning in medical image analysis and multimodal learning for clinical decision support*. Springer; 2018:3-11.
19. Baumgardner JR, Frederickson PO. Icosahedral discretization of the two-sphere. *SIAM J Num Anal*. 1985;22(6):1107-1115. doi:10.1137/0722066
20. Lyu I, Kang H, Woodward ND, Styner MA, Landman BA. Hierarchical spherical deformation for cortical surface registration. *Med Image Anal*. 2019;57:72-88. doi:10.1016/j.media.2019.06.013
21. Lyu I, Bao S, Hao L, et al. Labeling lateral prefrontal sulci using spherical data augmentation and context-aware training. *NeuroImage*. 2021;229:117758. doi:10.1016/j.neuroimage.2021.117758
22. Koenderink JJ, Van Doorn AJ. Surface shape and curvature scales. *Image vis Comput*. 1992;10(8):557-564.
23. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174. doi:10.2307/2529310
24. Choi SM, Kim BC, Nam TS, et al. Midbrain atrophy in vascular Parkinsonism. *Eur Neurol*. 2011;65(5):296-301. doi:10.1159/000326907
25. Salvatore C, Cerasa A, Castiglioni I, et al. Machine learning on brain MRI data for differential diagnosis of Parkinson's disease and progressive supranuclear palsy. *J Neurosci Methods*. 2014;222:230-237.
26. Quattrone A, Nicoletti G, Messina D, et al. MR imaging index for differentiation of progressive supranuclear palsy from Parkinson disease and the Parkinson variant of multiple system atrophy. *Radiology*. 2008;246(1):214-221.
27. Lee H, Lee MJ, Kim EJ, Huh GY, Lee JH, Cho H. Iron accumulation in the oculomotor nerve of the progressive supranuclear palsy brain. *Sci Rep*. 2021;11(1):1-9.
28. Du G, Lewis MM, Styner M, et al. Combined R2* and diffusion tensor imaging changes in the substantia nigra in Parkinson's disease. *Mov Disord*. 2011;26(9):1627-1632. doi:10.1002/mds.23643
29. Langley J, Huddleston DE, Chen X, Sedlacik J, Zachariah N, Hu X. A multicontrast approach for comprehensive imaging of substantia nigra. *NeuroImage*. 2015;112:7-13. doi:10.1016/j.neuroimage.2015.02.045
30. Gaurav R, Yahia-Cherif L, Pyatigorskaya N, et al. Longitudinal changes in neuromelanin MRI signal in Parkinson's disease: a progression marker. *Mov Disord*. 2021;36(7):1592-1602.
31. Poston KL, Ua Cruadhlaoich MA, Santoso LF, et al. Substantia nigra volume dissociates Bradykinesia and rigidity from tremor in Parkinsons disease: a 7 Tesla imaging study. *J Parkinsons Dis*. 2020;10(2):591-604. doi:10.3233/JPD-191890
32. Lee H, Baek SY, Chun SY, Lee JH, Cho H. Specific visualization of neuromelanin-iron complex and ferric iron in the human post-mortem substantia nigra using MR relaxometry at 7T. *NeuroImage*. 2018;172:874-885.
33. Lee JH, Baek SY, Song Y, et al. The neuromelanin-related T 2* contrast in postmortem human substantia nigra with 7T MRI. *Sci Rep*. 2016;6(1):1-9.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Tessema AW, Lee H, Gong Y, et al. Automated volumetric determination of high R_2^* regions in substantia nigra: A feasibility study of quantifying substantia nigra atrophy in progressive supranuclear palsy. *NMR in Biomedicine*. 2022;e4795. doi:10.1002/nbm.4795